MONTHLY WEATHER REVIEW

Editor, W. J. HUMPHREYS

VOL. 63, No. 8
W.B. No. 1163

AUGUST 1935

CLOSED OCTOBER 3, 1935
ISSUED NOVEMBER 6, 1935

CORRELATION-PERIODOGRAM INVESTIGATION OF RAINFALL ON THE WESTERN COAST OF THE UNITED STATES

By ERVIN J. PROUSE[1]

There have been literally thousands of attempts to find cycles in meteorological phenomena, especially in rainfall and temperature. Among these is that of Alter, published in 1922, in which he demonstrated for rainfall the existence of a cycle of length equal to one-ninth that of the sun-spot cycle. In this work he examined data from each continent of the world and used the phase of the sun-spot variation, instead of the time, as independent variable.

In this work Alter examined exclusively the evidence regarding a definite cycle (namely, one-ninth that of sun-spots) and was not using the periodogram in a search for any other cycles which might exist. He has suggested to the writer that a search of Pacific coast rainfall be made by means of the correlation periodogram, to find whether additional periodicities or cycles exist. Incidentally, the Pacific coast data which he had used were from Oregon and California exclusively and, in the light of the present work, were dominated by the latter.

The data examined in this study are from all stations of long rainfall records in western parts of Washington, Oregon, and California. The data for the stations in California all extend from 1877 or earlier. For Washington and Oregon the corresponding date was 1879. In all cases the data extend to 1930.

Data of each of the stations were summed over thirds of a year as follows: January through April, May through August, and September through December. This division into thirds of a year was made so that the May to August 4 months' period of California data could be omitted. The rainfall in California for these months is very small and the resulting percentages of normal would be so variable that accidental variations might dominate the periodogram. The same method of division as for California was followed for Washington and Oregon. In the case of the Washington and Oregon data, this summer period was used.

To facilitate an analysis of data, it was desirable to determine over how large an area data could profitably be combined. Assuming that stations close together geographically would have high enough correlations to be represented together, a study was made of the correlations between the rainfall averages, by 4-month periods, for a number of local groupings of the stations; and as a result, the data were arranged in two groups for the periodogram calculation, as follows. Apparently, independent conditions are operating in these two adjoining sections of the Pacific Coast region:

Group I contains the following stations in Washington and Oregon:

WASHINGTON	OREGON—continued
North Head	Astoria
Port Townsend	Portland
Vancouver	Roseburg
	Cascade Locks
OREGON	The Dalles
Ashland	Yreka, Calif.
Albany	

Group II contains the following stations in California:

Salinas	Ventura
Stockton	Santa Barbara
Merced	Newhall
Santa Cruz	Tustin
Hollister	San Luis Obispo
San Jose	San Bernardino
Healdsburg	San Diego
Monterey	Los Angeles
Oakdale	

The combined percentages of normal rainfall for each of the two groups are given as table 1. The division of the data into thirds of a year was retained, and the percentage of normal for each of these thirds given with the exception of the middle third of each year for the California data.

The data of table 1 were analyzed by the correlation periodogram method. This method has been exhibited in a number of calculations by Alter and others. The graphical representation of the periodogram from California follows as figure 1 and of Washington-Oregon as figure 2.

In the case of the periodogram of California only 33 values were computed, since by omitting the middle period of 4 months the number of possible pairs of products was cut in half and consequently the number of values that it is desirable to compute was limited proportionally.

We find in studying the periodogram of California that the largest positive ratio of the 33 values was only 1.31, and that including negative values there were only six greater than 1. It would not be at all surprising to find, using random data, a chance value as large as the

[1] Thesis submitted to the University of Kansas in partial fulfillment of the requirements for the degree of Master of Arts, May 1933.

21683—35——1

greatest of these. For this reason, it would be entirely unsafe to draw any positive conclusions from the periodogram. We might suggest that a satisfactory analysis of the rainfall of California by methods similar to the above must wait until more data can be accumulated. However, for the Washington-Oregon group, where there is sufficient rainfall to use the middle third of the year, we have approximately 50 percent more products and can make a more critical examination.

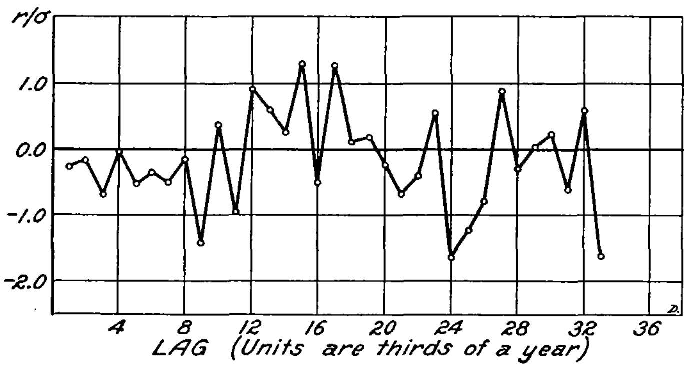

FIGURE 1.—Correlation periodogram of California rainfall.

We find among the 76 ratios in the periodogram of this group four positive values larger than 2, i. e., 2.48, 2.50, 2.19, 2.30, and one negative value greater than minus 2, i. e., −2.52. There are 15 positive values greater than 1, and 14 negative values greater than minus 1. This differs widely from the distribution we would expect by accident. We would expect only 1 positive value out of 70 to be as large as the smallest of the 4, and only 1 out of every 156 to be as large as the largest. We can be quite confident, therefore, because of the multiplicity of large values, that something nonaccidental is indicated in the sequence of rainfall values.

Even more striking than this is the fact that peaks follow each other at quite regular intervals. In a correlation periodogram any periodicity of length l that exists continuously throughout the data used will be manifested in the periodogram by peaks at l, $2l$, $3l$, etc.

A cosine curve of period equal to 4.2 datum intervals and with a phase of zero degrees at lag zero was superimposed on the graphical representation of the periodogram. This period was chosen merely as the best to emphasize the regular recurrence of peaks. The resemblance of the two curves as shown in figure 2 is very striking to the eye. The correlation between the periodogram and the cosine curve was computed and a value of $r=+0.44$ secured. For the range of the data, 76 values omitting lag 0, a sine curve would reach its maximum positive value 18 times. For all of the 18 maxima of the sine curve there was a periodogram peak within one datum interval on each side. On the other hand, only two positive peaks appeared in the periodogram which were not called for by the cosine curve, i. e., by the cycle of 1.4 years. In the case of the minimum values of the cosine curve we find agreement within one datum interval on each side in 16 out of 18 cases, and peaks appearing in the periodogram in but two cases where depressions were called for.

Inspection shows also in the periodogram a much longer swing of the values through a period of approximately 51 datum intervals. Twelve times the 4.2, which we have already considered, is 50.4 of such intervals. A least squares solution was made to find the two cosine terms of lengths 4.2 and 50.4 with phase zero at lag zero that will represent best the periodogram. The results are shown as figure 3. The amplitudes are $r=0.0685$ for the 4.2 and $r=0.626$ for the 50.4. The correlation between the sum of these terms and the original periodogram is $r=+0.612$ and $r/\sigma_r=+5.30$. Only 1 out of 30 million choices should have such a high ratio by accident. Since we have been limited to comparatively few choices there is no question that the pattern of the periodogram is nonaccidental. We have, therefore, a second independent criterion indicating a nonaccidental character for the results obtained.

Alter in working with rainfall data over many areas of the world, showed that a period related to the sunspot period and equal in length to one-ninth of that period existed. The cycle found in this work, 16.8 months, is almost exactly one-eighth the sunspot period. This is rather surprising since in Alter's data, which were dominated by California, the ninth harmonic showed as strongly as he found it any place in the world.

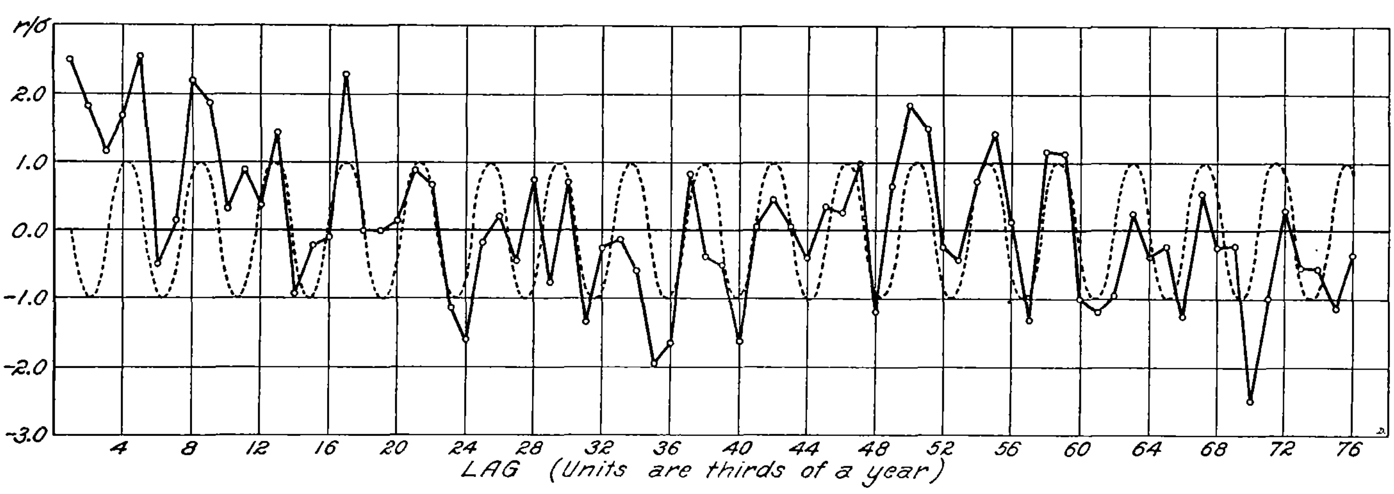

FIGURE 2.—Correlation periodogram of Oregon-Washington rainfall, with a 16.8-month sine curve superposed.

In order to check further the reality of the eighth harmonic, tables were formed from the first and second halves of the data using monthly values instead of the data as given in table 1.

The cycle exhibited is 16.8 months in length, therefore the first 17 values were written in a row; the next 17 in a row immediately under these, etc., so that the 1st, 18th, 35th, etc., were in a column, the 2d, 19th, 36th, etc., in a second column, and so on for 17 columns. Since the period is not exactly equal to 17 months there is a slowly increasing error in the positions of the data in the table. This is rectified by repeating the datum at the end of every fifth row.

In such tables the cycle investigated is not averaged out of column means, but all other variations are thus canceled more and more completely as the number of rows increases. The cycle will therefore stand out more prominently in the table means than in the individual data. The means of the tables from the halves of the data may be compared to learn whether they correlate in shape, amplitude of variation and phase. The resulting correlation between the two halves was $r=+0.33$, a value less than was expected from the appearance of the periodogram.

If we assume that this 16.8 month cycle is actually the eighth harmonic of the sunspot cycle, we will expect it to show the same sort of phenomena that Alter found for the ninth harmonic, that is, we should expect the halves to correlate better if we should follow his scheme and use sunspot phase instead of time as the independent variable. The method followed in making the adjustment was published in a paper that is not generally available.

If we plotted abscissae and ordinates on the same scale, these average values would form squares bounded by ordinates through the dates which limit them. The area between the axis of abscissae and the unknown curve, described above representing the actual value of the period at all times, would in the interval between two maxima or two minima necessarily equal the corresponding known square. Since these squares overlap, we know the value of a series of overlapping definite integrals of the unknown curve. From these data it is possible, assuming the simplest curve to be the true one, by the aid of a planimeter, to construct the curve without knowledge of its mathematical form. In doing this it is easier to choose some convenient period as the axis of abscissae and to measure departures from this period. Changing the axis in this way merely changes all the integrals by a known constant amount and changes the known squares into known rectangles. It is also practical to magnify the scale or ordinates very much over the scale of abscissae. Locating the curve consists first in measuring the area of each of the rectangles, then penciling in what appears to be the curve, measuring the definite integrals of the approximate curve with the planimeter, erasing for a new approximation, and repeating many times. In the curve of the sun-spot values reproduced as figure I, I have erased each part of the curve probably a hundred times. Although very laborious, the process, with enough patience, yields very good results. The accuracy of the period curve depends upon the accuracy with which the epochs of maxima and minima are obtained. A steep but narrow peak, such as that of 1861, may be unreal for this reason. However, due to the short duration of such a peak and the fact that it must almost immediately be counterbalanced, it will usually have little effect in an examination of data extending over many years.

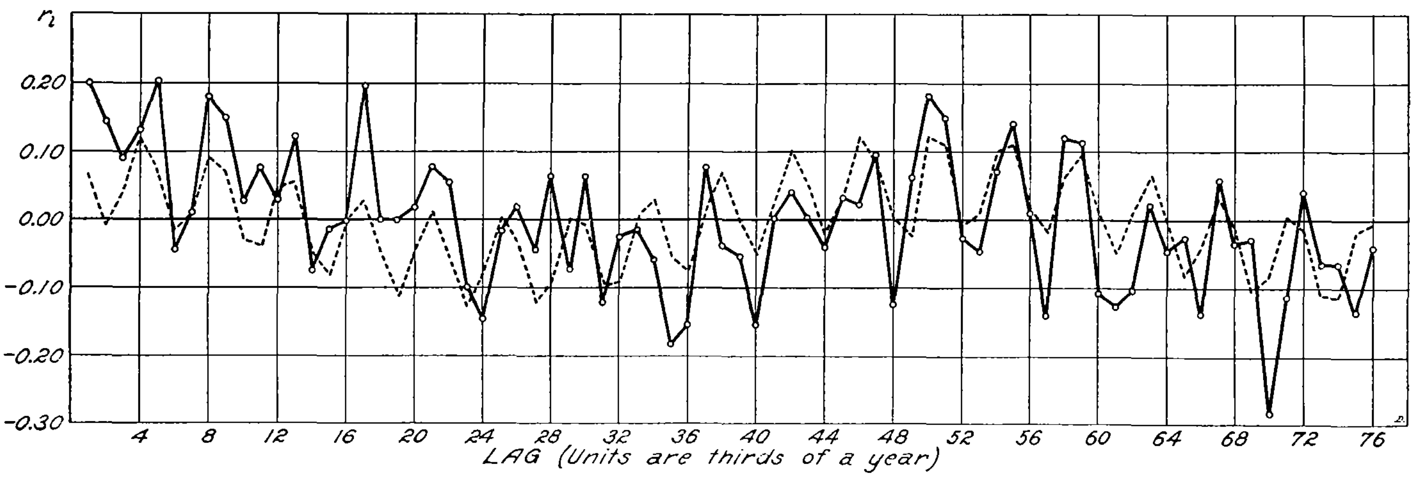

FIGURE 3.—Correlation periodogram of Oregon-Washington rainfall with combined 1.40- and 16.80-year sine curves superposed.

It is, therefore, thought best to quote it here. Tables and figures mentioned are of his paper (1). He says:

Let us consider the following problem; minima occurred in 1889, August, and in 1901, September. The next minimum occurred in 1913, May. It is evident that the sunspot period between the minima named above had values of 145 and 141 months, respectively. Let us examine the two maxima occurring between these dates. One occurred in 1894, February, and the other in 1906, May, with an interval of 147 months. This must have been the average value of the sunspot period between these dates. It is longer than the period obtained from either pair of minima named above, yet it occurs as part of each of them and contains no part that is not in one or the other of them. We are forced, therefore, to the conclusion that if continuous—

The length of the sunspot period is continuously varying and a value of the period obtained between successive maxima or successive minima is merely an average of all values passed through in this interval.

If we had a curve with time plotted along the axis of abscissae and the corresponding values of the sunspot period as ordinates, the average value of the sunspot period between 2 maxima or 2 minima occurring at t_1 and t_2 would be given by—

$$t_1-t_2 \text{ Average value } \int_{t_1}^{t_2} \frac{curve}{t_1-t_2}$$

In the preceding paragraph I have spoken of the sun-spot period at any date as a varying quantity, not even approximately constant through a single cycle. This may necessitate a definition of "period" somewhat different from what is ordinarily understood. I therefore give the following definition which will be adhered to whether referring to sun-spots or rainfall.

The length of the period at any date is the reciprocal of the rate of change of phase at that date and need not continue even approximately through a complete cycle.

From this curve I have taken the mean value of the sun-spot period for each year. These values are given as column 2 of table 2. Column 3 gives the departure from 15 months of one-ninth these values. Obviously, 15 months was chosen because it is the nearest integral number of months to one-ninth of a period. If, for example, the number given for any year in column 3 were plus nine, it would mean that during that year one-ninth of the sun-spot period was 16 months. If it were minus nine, it would mean that the period was 14 months. In the first case it would be necessary, working on a 15-month basis, to skip a month every 16 months as long as that length of period persisted, in the second case to repeat 1 every 14 months. We can thus construct a table of months to be repeated in the analysis of our rainfall data when the ninth of the sun-spot period is less than 15 months, or to be skipped (or better still, averaged with the next adjacent one) when the ninth is more than 15, in order that Wolfer's sun-spot maxima may all fall in one phase and his sun-spot minima in one.

Alter's table of months to be averaged or repeated in order to make this adjustment is reproduced as table 2 with additional months added to bring it up to date.

The correlation computed after adjustment was made was $r=+0.46$. This is much higher than the $+0.33$ correlation of the chronological data, confirming the evidence not only as to the reality of this cycle, but that it is actually related to the sun-spot cycle and equal to one-eighth of that cycle.

The writer examined Alter's California and Oregon data for the eighth harmonic and found a small correlation of $+0.178$ between the halves of the data. This appears probably due only to the few Oregon stations included among the California ones. He also examined the above data of Oregon and Washington for the ninth harmonic and found the negligible correlation of $+0.036$. There seems to exist very definitely in Washington and Oregon a different cycle from that which exists just as definitely in California and in many other parts of the world.

TABLE 1

Group	I	II	Group	I	II	Group	I	II
1877 a	-----	54	1895 a	89	116	1913 a	81	77
b	-----	x	b	147	x	b	141	x
c	-----	77	c	119	56	c	90	140
1878 a	-----	186	1896 a	151	101	1914 a	120	149
b	-----	x	b	149	x	b	71	x
c	-----	56	c	182	39	c	87	103
1879 a	120	92	1897 a	145	98	1915 a	81	141
b	179	x	b	91	x	b	103	x
c	96	144	c	130	52	c	118	82
1880 a	123	103	1898 a	86	34	1916 a	132	147
b	115	x	b	104	x	b	131	x
c	80	160	c	93	60	c	70	129
1881 a	144	74	1899 a	123	87	1917 a	90	72
b	122	x	b	146	x	b	73	x
c	105	66	c	118	128	c	92	24
1882 a	107	87	1900 a	74	50	1918 a	88	104
b	75	x	b	133	x	b	69	x
c	123	63	c	93	119	c	81	123
1883 a	99	71	1901 a	100	100	1919 a	128	85
b	46	x	b	68	x	b	59	x
c	87	64	c	69	65	c	83	77
1884 a	77	207	1902 a	105	105	1920 a	65	75
b	104	x	b	84	x	b	93	x
c	92	120	c	120	102	c	130	144
1885 a	72	39	1903 a	84	107	1921 a	121	83
b	110	x	b	98	x	b	78	x
c	114	176	c	96	59	c	110	168
1886 a	95	120	1904 a	140	110	1922 a	78	54
b	95	x	b	51	x	b	78	x
c	110	45	c	95	114	c	89	162
1887 a	144	96	1905 a	69	132	1923 a	82	61
b	93	x	b	104	x	b	100	x
c	105	77	c	87	58	c	71	46
1888 a	90	100	1906 a	84	143	1924 a	63	57
b	157	x	b	116	x	b	46	x
c	76	153	c	121	137	c	115	110
1889 a	64	76	1907 a	102	151	1925 a	95	87
b	104	x	b	87	x	b	76	x
c	92	341	c	101	88	c	77	64
1890 a	135	114	1908 a	84	83	1926 a	73	157
b	97	x	b	116	x	b	113	x
c	51	65	c	70	72	c	114	136
1891 a	96	99	1909 a	106	177	1927 a	105	124
b	127	x	b	84	x	b	82	x
c	140	82	c	115	166	c	106	138
1892 a	77	72	1910 a	106	64	1928 a	97	67
b	99	x	b	76	x	b	40	x
c	115	160	c	107	42	c	87	112
1893 a	118	124	1911 a	74	171	1929 a	66	60
b	112	x	b	102	x	b	71	x
c	134	79	c	84	49	c	57	8
1894 a	147	62	1912 a	101	86	1930 a	-----	-----
b	106	x	b	174	45			
c	104	158	c	100	45			

TABLE 2.—*Data repeated or averaged in keeping rainfall periodicity in step with sun spots*

Averaged	Repeated	Averaged
1861 Mar., Sept. 1862 June 1863 June	1865 July 1866 July 1867 Mar., June, Sept., Dec. 1868 Jan., Apr., June, Aug., Nov. 1869 Feb., June, Oct. 1870 April, Oct. 1871 April	1872 April 1873 Sept. 1874 April, Sept. 1875 Mar., June, Nov. 1876 Feb., May, Aug., Nov. 1877 Jan., Apr., July, Sept., Dec. 1878 Mar., June, Aug., Nov. 1879 Mar., July, Nov. 1880 Apr., Oct. 1881 July 1883 Mar.

Repeated	Averaged	Repeated
1884 Jan., Sept. 1885 April, Oct. 1886 Jan., May, Sept. 1887 Jan., May, Sept. 1888 Jan., May, Sept. 1889 Feb.	1891 Jan. 1894 May 1895 Jan., Sept. 1896 April 1897 Mar. 1898 Jan., Dec. 1899 Dec. 1901 Jan., Nov. 1902 June 1903 Sept. 1909 July 1913 Jan.	1915 Jan. 1917 July.

Repeat:	Average:
1915 Jan. 1917 July 1918 June 1919 Feb., Sept. 1920 Mar., Sept. 1921 Mar., Oct. 1922 May, Dec. 1923 July	1926 Jan. 1927 Jan. 1928 Jan.

REFERENCES

(1) Dinsmore Alter. A rainfall period equal to one-ninth the sunspot period. Kansas University Science Bulletin, vol. xiii, No. 11, July 1922.

(2) G. Udney Yule. Presidential address, Sec. V. Jour. Roy. Stat. Soc., January 1926.

(3) Dinsmore Alter. A group or correlation periodogram. MONTHLY WEATHER REVIEW, June 1927.

RELATION OF THE EXTREMES OF NORMAL DAILY TEMPERATURE TO THE SOLSTICES

By Edward H. Bowie

[Weather Bureau, San Francisco, Calif., August 1935]

A true normal daily temperature is defined as one that has been computed from a long series of values of hourly temperatures for each day, derived from automatically-recording thermometers.[1] There are numerous records of this character that cover periods of upwards of 20 years at many Weather Bureau stations; but these, according to Marvin and Day, are insufficient in number adequately to represent the details of the climatic conditions over an area the size of the United States.

Moreover, the labor of computing normals from hourly readings is too great to justify their general preparation. In lieu thereof, normal daily temperatures, based on the maxima and minima of temperature, have been computed, since they are nearly the same as the normal daily temperatures determined from hourly readings for similar periods of time. Such normals are given in Supplement

[1] MONTHLY WEATHER REVIEW, Supplement No. 25, Normals of Daily Temperature for the United States, by Marvin and Day.